Research Statement

Alexander A. Sherstov

June 2015

Can machines provably learn from past experience?

Are quantum computers substantially more powerful than classical ones?

What are the computational limits of a streaming algorithm? A neural network? A computer chip?

At first glance, this looks like a hodgepodge of unrelated questions. Surprisingly, the research community has discovered over the past forty years that these questions are closely related. They can all be distilled to a single mathematical concept, *communication*, and they can all be fruitfully studied within the framework of a single mathematical theory, *communication complexity theory*. The aim of my research is to understand communication as a pervasive computational phenomenon and use my findings to answer fundamental questions in computational learning, quantum computing, and other models of computation. I have solved several longstanding open problems in these areas, which I describe below along with my vision for future research.

1 Background

Communication is a common bottleneck. It arises whenever a computational problem requires coordination among several parties, each with an incomplete view of the input. Communication complexity theory, initiated in the seminal work of A. Yao [61] forty years ago, is an area of theoretical computer science that studies communication as a computational resource, measured in bits. The simplest setting features two communicating parties, call them Alice and Bob, and a function f with two arguments. Alice has an input x, Bob has an input y, and their objective is to compute $f(x, y)$ by communicating back and forth according to an agreed-upon set of rules, called a *communication protocol*. Communication complexity theory studies the minimum amount of communication required to compute f, a quantity known as the *communication complexity* of f. A straightforward but prohibitively expensive communication protocol is for Alice to send her entire input to Bob. For some functions, this brute force approach is provably optimal, whereas for others one can accomplish the task with surprisingly little communication. To cite a classical example, if Alice and Bob each hold an n-bit string, they can test their strings for equality with accuracy 99% by exchanging only eight bits, no matter how large n is. (Think of verifying the equality of two large, geographically separated datasets!) From the simple two-party model just described to nondeterministic and quantum communication to sophisticated multiparty systems with overlapping inputs, communication complexity theory studies a wide variety of models and does so in a clean, abstract, and unified way.

Surprisingly, the communication bottleneck is often present even when no transmission of data occurs. Indeed, many computational problems involve an implicit flow of information among several components and are therefore subject to the limitations of multiparty communication. Over the years, communication complexity theory has proven to be a powerful tool in virtually all areas of theoretical computer science and beyond, including computational learning, pseudorandom generators, streaming algorithms, quantum computing, data structures, mechanism design, and chip layout. In what follows, I describe my contributions to communication complexity theory itself (Section 2) as well as computational learning (Section 3) and quantum computing (Section 4), concluding with my program for future work (Section 5).

2 Communication Complexity

My contributions in communication complexity theory include: (1) a powerful and general method for estimating communication complexity, the *pattern matrix method*; (2) determination of the communication complexity of the multiparty problem known as *set disjointness*, a central problem in the area; and (3) a number of results establishing the relative power of different models of communication. My work solves a 22-year-old problem due to Babai, Frankl, and Simon, a 16-year-old problem due to Wigderson, an 11-year-old problem due to Kushilevitz and Nisan, and a 10-year-old problem due to Krause and Pudlák.

The pattern matrix method. Research in communication complexity theory centers around proving that a communication problem $f(x, y)$ of interest admits no efficient protocol. In other words, Alice and Bob cannot compute $f(x, y)$ reliably unless their protocol communicates many bits. Proving such *lower bounds* on communication is difficult—after all, one has to somehow rule out every possible low-cost protocol. In [44, 47], I developed a novel method for proving communication lower bounds, the *pattern matrix method,* based on matrix analysis and linear programming duality. My method yields strong lower bounds on communication in a variety of models, including both classical and quantum communication channels, both high-accuracy communication protocols and protocols with accuracy vanishingly close to random guessing. To apply the pattern matrix method, it suffices to prove that the corresponding Boolean function cannot be approximated by a low-degree real polynomial. Since the approximation of Boolean functions by real polynomials is an extensively studied subject, my method immediately gives lower bounds for a broad new class of communication problems.

As a first application of the pattern matrix method (we will see several others), I solved a problem in circuit complexity due to Krause and Pudlák [32] that had been open for 10 years. The problem concerns two basic computational models: constant-depth circuits of $\wedge, \vee, \neg$ gates and constant-depth circuits of majority gates. Allender's classic result [2] states that every circuit of the first type can be efficiently converted into a depth-3 circuit of the second type. Krause and Pudlák asked if depth 2 suffices. I answered this question in the negative [44], proving that any conversion to depth 2 will involve an exponential blowup in circuit size. Thus, Allender's result from 1989 is best possible.

I am excited to report that the pattern matrix method has enabled considerable progress by other researchers on longstanding open problems, e.g., [16, 36, 17, 18, 19, 10, 21].

The set disjointness problem. Consider k parties trying to pick a time to meet. There are a total of n time slots, and each party's availability is a subset of these n time slots. A given party knows his own availability and perhaps the availability of many others—but not all. The goal is to determine with high probability (say, 99%) whether there is a time slot that works for

everyone. This is the most studied problem in communication complexity theory, known as the *set disjointness problem*. Abstractly, the set disjointness problem is for k parties to determine whether the given k subsets $S_1, S_2, \ldots, S_k \subseteq \{1, 2, \ldots, n\}$ have nonempty intersection, where no party knows all the k subsets.

In a striking 1994 paper, Grolmusz [22] proved that the problem can be solved using roughly $O(n/2^k)$ bits of communication. This was known to be optimal for $k = 2$ parties [24, 40, 9], but starting at $k = 3$ the optimality of Grolmusz's communication protocol remained open for over a decade. This state of affairs changed drastically several weeks after the publication of my pattern matrix paper, with an exciting flurry of papers [36, 17, 10] using my method to prove a lower bound of roughly $\Omega(n/2^{k^3})^{1/(k+1)}$ bits on the communication required. This result is strong for small constant k but deteriorates rapidly as k grows. Yet to match Grolmusz's result, one would need to prove that the communication requirements remain high for as many as $k = \Theta(\log n)$ parties! Once again, progress on the set disjointness problem had come to a stall.

Four years later, I resolved the problem in its entirety [48, 55]. Specifically, I proved a lower bound of $\Omega(\sqrt{n}/2^k k)$ on the communication requirements of k-party set disjointness. It had been an open problem for 16 years to prove this lower bound even in the special case $k = 4$; see Wigderson [7]. My result is optimal for quantum communication protocols, and by Grolmusz it is within a square of optimal for classical protocols.

Unbounded-error communication. One typically requires the communication protocol to output the correct answer with high probability, say 99%. A notable exception is the *unbounded-error model*, where it is enough to produce the correct answer with any probability above 50% (say, 50% plus an exponentially small amount). This unusual model turns out to be one of the richest mathematically and most significant practically, with applications ranging from matrix analysis to high-dimensional geometry and machine learning.

In a joint work [42], Razborov and I solved an open problem on unbounded-error communication that had been open for 22 years. This problem, due to Babai, Frankl, and Simon [6], asks whether unbounded-error communication is more powerful than the *polynomial hierarchy* (PH), another fundamental model which generalizes the familiar notion of nondeterminism. We gave a negative answer to this question, exhibiting a function from the polynomial hierarchy with almost maximum complexity in the unbounded-error model. As I discuss below, our result explains why no efficient algorithm has been discovered for learning formulas in disjunctive normal form, a famous problem which in 1984 gave rise to computational learning theory and has since remained open.

Product and nonproduct distributions. I have said a lot about *why* proving lower bounds on communication is important. But *how* does one go about proving them, at least in the two-party setting? The basis for most proofs is Yao's minimax theorem [62], which states: to prove that a communication problem $f(x, y)$ is hard, it is necessary and sufficient to find a probability distribution on the (x, y) pairs under which any low-cost protocol fails to reliably tell apart the cases $f(x, y) = 0$ and $f(x, y) = 1$. Much of the art and magic of communication complexity theory is constructing such probability distributions. The simplest distributions to work with are *product distributions*, whereby Alice and Bob's arguments x and y are distributed independently. Kushilevitz and Nisan [33] asked whether the use of product distributions always leads to near-optimal results. This methodological question remained open for 11 years. I resolved in [45], with an answer in the negative. Specifically, I proved the existence of a Boolean function $f\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ whose communication complexity is the highest possible, $\Omega(n)$, but for which a product distribution can only certify a trivial $\Omega(1)$ lower bound.

In an earlier effort [43] toward solving the Kushilevitz-Nisan problem, I studied *discrepancy*, one

of the most important complexity measures of a communication problem. I constructed communication problems with an exponential gap between the discrepancy under product vs. nonproduct distributions. My work additionally proved that unbounded-error communication protocols can be exponentially more efficient than a related communication model, PP^{cc}—more than 20 years after the introduction of the two models. I received the "Best Student Paper" award for these contributions at the Twenty-Second IEEE Conference on Computational Complexity (CCC 2007). Independently of my work, Buhrman, Vereshchagin, and de Wolf [13] obtained the same separation for PP^{cc} with much different techniques.

Approximation by polynomials. Approximation theory has been a vital source of inspiration in my contributions to communication complexity theory, including the pattern matrix method and set disjointness problem. Conversely, the insights that I gained along the way have allowed me to resolve several questions in polynomial approximation. One such, due to Buhrman et al. [12], is whether every bounded polynomial on the Boolean hypercube can be made robust to noise in the inputs with only a constant-factor increase in degree. Such robustness to noise becomes crucial if one needs to do anything nontrivial with the polynomials, e.g., compose them. The problem had been unsolved for 9 years when I gave a complete solution to it in [52]. For any $\delta > 0$ and any polynomial $p\colon \{0,1\}^n \to [-1,1]$, I showed how to construct a corresponding polynomial $p_{\text{robust}}\colon \mathbb{R}^n \to \mathbb{R}$ of degree only $O(\deg p + \log \frac{1}{\delta})$ that is robust to noise in the inputs: $|p(x) - p_{\text{robust}}(x+\epsilon)| < \delta$ for all Boolean inputs $x \in \{0,1\}^n$ and their perturbations $\epsilon \in [-1/3, 1/3]^n$. This result is optimal with respect to all parameters.

Another well-known open problem was to determine the degree of a real polynomial required to approximate the function $f(x) = \bigwedge_{i=1}^{n} \bigvee_{j=1}^{m} x_{i,j}$ to within a small constant on every point, with $\Theta(\sqrt{nm})$ being the conjectured answer. This simple-sounding problem evaded complexity theorists and quantum researchers for 19 years [38, 57, 23, 3, 51] and was recently re-posed as an open problem by Aaronson [1]. My paper [50] gives a complete solution to this problem, showing that a polynomial of degree $\Theta(\sqrt{nm})$ is required. The same result was obtained independently by Bun and Thaler [14], using related techniques.

3 Computational Learning Theory

Humans have a striking ability to learn from past experience. Is it possible to replicate this ability with a computer algorithm, at least for well-defined and highly structured tasks? There is an increasing need for such algorithms, from computer vision to electronic commerce to social networking. *Computational learning theory* studies precise mathematical models of learning and aims to design efficient learning algorithms or to understand why some learning tasks are inherently hard. In the canonical model due to Valiant [59], known as PAC learning, the learner receives values of an unknown Boolean function f at independently sampled points from the domain. The goal is to efficiently find an accurate approximation to f, based on these training examples and the fact that f belongs to some known class $\mathscr{C}$ of functions.

Efficient algorithms have been discovered for several classes $\mathscr{C}$. Other important classes, such as DNF formulas, constant-depth circuits, and intersections of halfspaces, have remained off-limits despite decades of extensive research. My contribution was to establish strong intractability results for these learning problems, thereby ruling out a broad range of algorithmic approaches.

Intersections of halfspaces. A *halfspace* in $\mathbb{R}^n$, also known as a *linear threshold function*, is a Boolean function of the form $f(x) = \operatorname{sgn}(\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n - \theta)$ for some fixed reals $\alpha_1, \alpha_2, \ldots, \alpha_n, \theta$. While there are many efficient algorithms for learning a single halfspace,

it is a longstanding challenge in the area to efficiently learn intersections of two or more halfspaces [11, 34, 60, 5, 28, 29, 27]. Klivans and I explained this lack of progress by establishing near-optimal, exponential lower bounds [30] on the sample complexity of learning the intersection of n halfspaces in the *statistical-query model* [25]. This work received the "Best Student Paper" award at the Nineteenth Annual Conference on Learning Theory (COLT 2006). In subsequent research [31], we ruled out an efficient learning algorithm for intersections of n halfspaces even in the more powerful PAC model, under widely accepted cryptographic assumptions. Our intractability results closed an important line of research in computational learning.

The only possibility left open by our work was that of efficiently learning the intersection of k halfspaces for small k, such as $k = 2$. I addressed this question in [51, 53]. I focused on learning *by sign-representing polynomials*, a generalization of the perceptron algorithm and one of the few methods known for learning under arbitrary distributions. I proved that the intersection of even two halfspaces requires a sign-representing polynomial of the largest possible degree, $\Theta(n)$, meaning that this approach cannot yield anything better than a trivial, exponential-time algorithm. This work solved an open problem due to Klivans [26] and received the "Best Student Paper" award at the Fiftieth Annual Symposium on Foundations of Computer Science (FOCS 2009).

DNF formulas. Progress in computational learning theory has been achieved in great part via *analytic methods*, such as representations of unknown functions by halfspaces or their higher-degree analogues. I have contributed several articles on analytic complexity measures of learning problems, clarifying the relations among them and exposing the limitations of analytic methods. In [43], I proved an exponential separation between the *sign-rank* and *margin complexity* of a natural learning problem, showing that high-margin classification can be off-limits even on learning problems with a simple geometry. In a follow-up work [42], Razborov and I studied *disjunctive normal form (DNF) formulas*, which are expressions such as $x_1 \vee (x_2 \wedge \overline{x}_7 \wedge x_5) \vee (x_3 \wedge \overline{x}_5)$. Learning DNF formulas efficiently is a central unresolved problem in computational learning theory. We proved that DNF formulas of size n have sign-rank exponential in n, i.e., a very complicated geometry. Our work gives the first unconditional, exponential lower bound for learning DNF formulas by a wide range of algorithms, including perceptrons of arbitrary degree and any other finite-dimensional kernel methods.

The Minsky–Papert barrier. DNF formulas can be thought of as a special kind of Boolean circuits, with depth bounded by 2. Relaxing the depth bound to an arbitrary constant, such as 29, results in one of the most studied classes in theoretical computer science, referred to fittingly as *constant-depth circuits.* A basic question that has long puzzled researchers in learning theory and communication complexity theory alike is whether every constant-depth polynomial-size circuit has a *perceptron* of relatively low degree, i.e., can be represented by the sign of a low-degree real polynomial. An answer in the affirmative would give an efficient learning algorithm for this concept class; a strong answer in the negative would be an important impossibility result in communication complexity theory.

The question turned out to be surprisingly difficult. In their seminal monograph 45 years ago, Minsky and Papert [37] constructed a constant-depth polynomial-size circuit in n variables that requires a sign-representing polynomial of degree $\Omega(n^{1/3})$. The only subsequent improvement, 12 years ago, was due to O'Donnell and Servedio [39], who obtained a bound of $\Omega(n^{1/3} \log^c n)$ for any constant $c > 0$. They formally posed the problem of breaking Minsky and Papert's "1/3" barrier and obtaining a lower bound with a larger exponent.

My recent work [54] gives a detailed solution to this problem. For any depth d, I construct a size-n depth-d circuit (in fact, a tree circuit, also known as a *formula*) that requires a sign-representing polynomial of degree $\Omega(n^{(d-1)/(2d-1)})$. For d large, this lower bound nearly matches the famous $O(\sqrt{n})$ upper bound for arbitrary formulas [39, 20, 4, 35]. My result proves a decade-old conjecture due to O'Donnell and Servedio [39] and breaks the longstanding Minsky-Papert barrier. In a recent follow-up work [56], I obtain a polynomially stronger lower bound of $\Omega(\sqrt{n})$ for an explicitly given constant-depth circuit, resolving a question due to Bun and Thaler [15].

4 Quantum Communication

Today's fastest algorithms for factoring integers run in exponential time. Indeed, the security of the widely used RSA public-key cryptosystem hinges on our inability (as yet?) to factor integers efficiently. In 1994, P. Shor made a spectacular discovery [58]: he developed a polynomial-time algorithm for factoring integers on a *quantum* computer. Shor's breakthrough spurred an intense effort to understand the power and limitations of quantum computing. In this line of research, communication remains a natural and important computational resource to study. Since the introduction of quantum communication complexity by Yao [63] twenty-two years ago, the subject has become an integral part of communication complexity theory and is an essential part of my research program.

Lower bounds for quantum communication. My technique [44] for communication lower bounds, the pattern matrix method, applies not only to classical communication but to quantum as well. As of this writing, my method is one of the strongest and most general tools known for analyzing quantum communication complexity. In particular, my results broadly subsume the celebrated lower bounds due to Razborov [41], including the $\Omega(\sqrt{n})$ lower bound for the set disjointness problem. (Razborov's proof technique was entirely different). In my most recent work on the topic [46], I used the pattern matrix method to exhibit a new class of communication problems for which quantum protocols are essentially no better than their classical counterparts, broadly generalizing previous work.

Direct product theorems. If it takes R units of a resource (such as time, space, or communication) to solve a single instance of a problem, how much of the resource is necessary to solve n independent instances? Common sense suggests an answer in the ballpark of nR, since otherwise the algorithm seems to be forced to guess randomly, resulting in an exponentially small probability of successful computation. Such statements are called *direct product theorems*. Their seeming simplicity is deceptive. Proving direct product theorems is notoriously hard—and in some models actually impossible. In particular, it is a longstanding problem to prove a direct product theorem for quantum communication ("solving n instances of a communication problem requires $\Omega(n)$ times the communication cost of a single instance, even for probability of correctness exponentially close to random guessing"). Prior to my work, progress on this question was quite limited. My recent paper [49] comes close to fully resolving the question. More precisely, I prove a direct product theorem for *generalized discrepancy*, which from a practical standpoint is equivalent to quantum communication complexity: the latter is always at least as large as the former, and they are equivalent for any communication problem ever studied. In particular, my work gives a direct product theorem for every quantum communication problem studied in the literature.

5 Future Directions

I will focus here on three of the most important open problems in communication complexity theory, which are a top priority of my research program. All three are basic scientific questions about the power and limitations of communication in the physical universe that we occupy. Their resolution will additionally have a substantial impact on computational learning theory, circuit complexity, quantum computing, and other areas.

Quantifiers in communication. A deep idea in computer science is the use of existential and universal quantifiers as a vehicle in computation. The familiar class P involves no quantifiers, whereas NP features a single quantifier (corresponding to the existence of a membership certificate). Allowing two or more quantifiers results in a class known as the *polynomial hierarchy*, denoted PH. Understanding the computational power of quantifiers is one of the most important open problems in modern science. Analogous to these Turing complexity classes, two-party communication has its own classes P, NP, PH. It is a central challenge in communication complexity theory to exhibit a communication problem not solvable in the polynomial hierarchy. I am developing a novel, spectral line of attack on the problem. If successful, my approach will also give a nonuniform circuit family of quasipolynomial size for PAC learning DNF formulas and constant-depth circuits under an arbitrary (but fixed) distribution, solving a major open problem in computational learning theory.

Multiparty communication. It is a significant challenge to analyze the communication requirements of functions whose arguments are distributed among many parties. This challenge gets harder as the number of parties grows. As usual, we are interested in the standard model of multiparty communication, which allows an arbitrary overlap among the arguments available to different parties. Known approaches in this setting break down completely as soon as the number of parties exceeds the *logarithm* of the bit length of the function's arguments, a famous obstacle in theoretical computer science known as the *logarithmic barrier* [8]. Breaking this barrier is of basic scientific importance, not to mention its vast consequences for circuit complexity theory and other areas. I am actively working on resolving this challenge.

Quantum vs. classical communication. Whether quantum communication can be substantially more powerful than classical is a major open problem in both quantum computing and communication complexity theory. Quantum protocols are known to be exponentially more efficient on a number of modified tasks involving communication (such as sampling or computing partial functions), but the original question remains wide open. It is closely related to what is perhaps the most famous open problem in communication, the *log-rank conjecture*: roughly speaking, both problems ask whether a *combinatorial view* of communication is equivalent to an *analytic view*.

References

[1] S. Aaronson. The polynomial method in quantum and classical computing. In *Proceedings of the Forty-Ninth Annual IEEE Symposium on Foundations of Computer Science* (FOCS), page 3, 2008.

[2] E. Allender. A note on the power of threshold circuits. In *Proceedings of the Thirtieth Annual IEEE Symposium on Foundations of Computer Science* (FOCS), pages 580–584, 1989.

[3] A. Ambainis. Polynomial degree and lower bounds in quantum complexity: Collision and element distinctness with small range. *Theory of Computing*, 1(1):37–46, 2005.

[4] A. Ambainis, A. M. Childs, B. Reichardt, R. Špalek, and S. Zhang. Any AND-OR formula of size N can be evaluated in time $N^{1/2+o(1)}$ on a quantum computer. In *Proceedings of the*

Forty-Eighth Annual IEEE Symposium on Foundations of Computer Science (FOCS), pages 363–372, 2007.

[5] R. I. Arriaga and S. Vempala. An algorithmic theory of learning: Robust concepts and random projection. *Mach. Learn.*, 63(2):161–182, 2006.

[6] L. Babai, P. Frankl, and J. Simon. Complexity classes in communication complexity theory. In *Proceedings of the Twenty-Seventh Annual IEEE Symposium on Foundations of Computer Science* (FOCS), pages 337–347, 1986.

[7] L. Babai, T. P. Hayes, and P. G. Kimmel. The cost of the missing bit: Communication complexity with help. *Combinatorica*, 21(4):455–488, 2001.

[8] L. Babai, N. Nisan, and M. Szegedy. Multiparty protocols, pseudorandom generators for logspace, and time-space trade-offs. *J. Comput. Syst. Sci.*, 45(2):204–232, 1992.

[9] Z. Bar-Yossef, T. S. Jayram, R. Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *J. Comput. Syst. Sci.*, 68(4):702–732, 2004.

[10] P. Beame and T. Huynh. Multiparty communication complexity and threshold circuit size of AC^0. *SIAM J. Comput.*, 41(3):484–518, 2012.

[11] A. Blum and R. Kannan. Learning an intersection of a constant number of halfspaces over a uniform distribution. *J. Comput. Syst. Sci.*, 54(2):371–380, 1997.

[12] H. Buhrman, I. Newman, H. Röhrig, and R. de Wolf. Robust polynomials and quantum algorithms. *Theory Comput. Syst.*, 40(4):379–395, 2007.

[13] H. Buhrman, N. K. Vereshchagin, and R. de Wolf. On computation and communication with small bias. In *Proceedings of the Twenty-Second Annual IEEE Conference on Computational Complexity* (CCC), pages 24–32, 2007.

[14] M. Bun and J. Thaler. Dual lower bounds for approximate degree and Markov-Bernstein inequalities. In *Proceedings of the Fortieth International Colloquium on Automata, Languages and Programming* (ICALP), pages 303–314, 2013.

[15] M. Bun and J. Thaler. Hardness amplification and the approximate degree of constant-depth circuits. In *Electronic Colloquium on Computational Complexity (ECCC)*, 2013. Report TR13-151.

[16] A. Chattopadhyay. Discrepancy and the power of bottom fan-in in depth-three circuits. In *Proceedings of the Forty-Eighth Annual IEEE Symposium on Foundations of Computer Science* (FOCS), pages 449–458, 2007.

[17] A. Chattopadhyay and A. Ada. Multiparty communication complexity of disjointness. In *Electronic Colloquium on Computational Complexity (ECCC)*, January 2008. Report TR08-002.

[18] M. David and T. Pitassi. Separating NOF communication complexity classes RP and NP. In *Electronic Colloquium on Computational Complexity (ECCC)*, February 2008. Report TR08-014.

[19] M. David, T. Pitassi, and E. Viola. Improved separations between nondeterministic and randomized multiparty communication. *ACM Transactions on Computation Theory (TOCT)*, 1(2), 2009.

[20] E. Farhi, J. Goldstone, and S. Gutmann. A quantum algorithm for the Hamiltonian NAND tree. *Theory of Computing*, 4(1):169–190, 2008.

[21] D. Gavinsky and A. A. Sherstov. A separation of NP and coNP in multiparty communication complexity. *Theory of Computing*, 6(10):227–245, 2010.

[22] V. Grolmusz. The BNS lower bound for multi-party protocols in nearly optimal. *Inf. Comput.*, 112(1):51–54, 1994.

[23] P. Høyer, M. Mosca, and R. de Wolf. Quantum search on bounded-error inputs. In *Proc. of the 30th International Colloquium on Automata, Languages, and Programming (ICALP)*,

pages 291–299, 2003.

[24] B. Kalyanasundaram and G. Schnitger. The probabilistic communication complexity of set intersection. *SIAM J. Discrete Math.*, 5(4):545–557, 1992.

[25] M. J. Kearns. Efficient noise-tolerant learning from statistical queries. *J. ACM*, 45(6):983–1006, 1998.

[26] A. R. Klivans. *A Complexity-Theoretic Approach to Learning*. PhD thesis, MIT, 2002.

[27] A. R. Klivans, P. M. Long, and A. K. Tang. Baum's algorithm learns intersections of halfspaces with respect to log-concave distributions. In *Proceedings of the Thirteenth International Workshop on Randomization and Computation* (RANDOM), pages 588–600, 2009.

[28] A. R. Klivans, R. O'Donnell, and R. A. Servedio. Learning intersections and thresholds of halfspaces. *J. Comput. Syst. Sci.*, 68(4):808–840, 2004.

[29] A. R. Klivans and R. A. Servedio. Learning intersections of halfspaces with a margin. *J. Comput. Syst. Sci.*, 74(1):35–48, 2008.

[30] A. R. Klivans and A. A. Sherstov. Unconditional lower bounds for learning intersections of halfspaces. *Machine Learning*, 69(2–3):97–114, 2007. Preliminary version in *Proceedings of the Nineteenth Annual Conference on Computational Learning Theory* (COLT), 2006.

[31] A. R. Klivans and A. A. Sherstov. Cryptographic hardness for learning intersections of halfspaces. *J. Comput. Syst. Sci.*, 75(1):2–12, 2009. Preliminary version in *Proceedings of the Forty-Seventh Annual IEEE Symposium on Foundations of Computer Science* (FOCS), 2006.

[32] M. Krause and P. Pudlák. On the computational power of depth-2 circuits with threshold and modulo gates. *Theor. Comput. Sci.*, 174(1–2):137–156, 1997.

[33] E. Kushilevitz and N. Nisan. *Communication complexity*. Cambridge University Press, 1997.

[34] S. Kwek and L. Pitt. PAC learning intersections of halfspaces with membership queries. *Algorithmica*, 22(1/2):53–75, 1998.

[35] T. Lee. A note on the sign degree of formulas, 2009. Available at `http://arxiv.org/abs/0909.4607`.

[36] T. Lee and A. Shraibman. Disjointness is hard in the multiparty number-on-the-forehead model. *Computational Complexity*, 18(2):309–336, 2009.

[37] M. L. Minsky and S. A. Papert. *Perceptrons: An Introduction to Computational Geometry*. MIT Press, Cambridge, Mass., 1969.

[38] N. Nisan and M. Szegedy. On the degree of Boolean functions as real polynomials. *Computational Complexity*, 4:301–313, 1994.

[39] R. O'Donnell and R. A. Servedio. New degree bounds for polynomial threshold functions. *Combinatorica*, 30(3):327–358, 2010.

[40] A. A. Razborov. On the distributional complexity of disjointness. *Theor. Comput. Sci.*, 106(2):385–390, 1992.

[41] A. A. Razborov. Quantum communication complexity of symmetric predicates. *Izvestiya of the Russian Academy of Sciences, Mathematics*, 67:145–159, 2002.

[42] A. A. Razborov and A. A. Sherstov. The sign-rank of AC^0. *SIAM J. Comput.*, 39(5):1833–1855, 2010. Preliminary version in *Proceedings of the Forty-Ninth Annual IEEE Symposium on Foundations of Computer Science* (FOCS), 2008.

[43] A. A. Sherstov. Halfspace matrices. *Computational Complexity*, 17(2):149–178, 2008. Preliminary version in *Proceedings of the Twenty-Second Annual IEEE Conference on Computational Complexity* (CCC), 2007.

[44] A. A. Sherstov. Separating AC^0 from depth-2 majority circuits. *SIAM J. Comput.*, 38(6):2113–2129, 2009. Preliminary version in *Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing* (STOC), 2007.

[45] A. A. Sherstov. Communication complexity under product and nonproduct distributions.

Computational Complexity, 19(1):135–150, 2010. Preliminary version in *Proceedings of the Twenty-Third Annual IEEE Conference on Computational Complexity* (CCC), 2008.

[46] A. A. Sherstov. On quantum-classical equivalence for composed communication problems. *Quantum Information & Computation*, 10(5-6):435–455, 2010.

[47] A. A. Sherstov. The pattern matrix method. *SIAM J. Comput.*, 40(6):1969–2000, 2011. Preliminary version in *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing* (STOC), 2008.

[48] A. A. Sherstov. The multiparty communication complexity of set disjointness. In *Proceedings of the Forty-Fourth Annual ACM Symposium on Theory of Computing* (STOC), pages 525–544, 2012.

[49] A. A. Sherstov. Strong direct product theorems for quantum communication and query complexity. *SIAM J. Comput.*, 41(5):1122–1165, 2012. Preliminary version in *Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing* (STOC), 2011.

[50] A. A. Sherstov. Approximating the AND-OR tree. *Theory of Computing*, 9(20):653–663, 2013.

[51] A. A. Sherstov. The intersection of two halfspaces has high threshold degree. *SIAM J. Comput.*, 42(6):2329–2374, 2013. Preliminary version in *Proceedings of the Fiftieth Annual IEEE Symposium on Foundations of Computer Science* (FOCS), 2009.

[52] A. A. Sherstov. Making polynomials robust to noise. *Theory of Computing*, 9:593–615, 2013. Preliminary version in *Proceedings of the Forty-Fourth Annual ACM Symposium on Theory of Computing* (STOC), 2012.

[53] A. A. Sherstov. Optimal bounds for sign-representing the intersection of two halfspaces by polynomials. *Combinatorica*, 33(1):73–96, 2013. Preliminary version in *Proceedings of the Forty-Second Annual ACM Symposium on Theory of Computing* (STOC), 2010.

[54] A. A. Sherstov. Breaking the Minsky-Papert barrier for constant-depth circuits. In *Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing* (STOC), pages 223–232, 2014. Full version available as ECCC Report TR14-009, January 2014.

[55] A. A. Sherstov. Communication lower bounds using directional derivatives. *J. ACM*, 61(6):1–71, 2014. Preliminary version in *Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing* (STOC), 2013.

[56] A. A. Sherstov. The power of asymmetry in constant-depth circuits. Manuscript, 2015.

[57] Y. Shi. Approximating linear restrictions of Boolean functions. Manuscript, 2002.

[58] P. W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Comput.*, 26(5):1484–1509, 1997.

[59] L. G. Valiant. A theory of the learnable. *Commun. ACM*, 27(11):1134–1142, 1984.

[60] S. Vempala. A random sampling based algorithm for learning the intersection of halfspaces. In *Proceedings of the Thirty-Eighth Annual IEEE Symposium on Foundations of Computer Science* (FOCS), pages 508–513, 1997.

[61] A. C.-C. Yao. Some complexity questions related to distributive computing. In *Proceedings of the Eleventh Annual ACM Symposium on Theory of Computing* (STOC), pages 209–213, 1979.

[62] A. C.-C. Yao. Lower bounds by probabilistic arguments. In *Proceedings of the Twenty-Fourth Annual IEEE Symposium on Foundations of Computer Science* (FOCS), pages 420–428, 1983.

[63] A. C.-C. Yao. Quantum circuit complexity. In *Proceedings of the Thirty-Fourth Annual IEEE Symposium on Foundations of Computer Science* (FOCS), pages 352–361, 1993.